

Streaming mit Apache Kafka

Ein Einstieg in das Stream Processing mit Kafka Streams DSL und KSQL

Agenda

- Über mich
- Die Welt der Daten
- Message Broker/Kafka Basics
- Stream Processing in der Theorie
- Showcase mit Kafka Streaming API und KSQL

Über mich

Über mich

Thomas Müller

- Senior Software-Developer im Bereich Java-Backend-Systeme bei diva-e
- Lieblings-Spielplätze: Apache Kafka, Hazelcast, Spring-Boot-Applikationen
- Email: thomas.mueller@diva-e.com
- Twitter: @zaroselectro



Die Welt der Daten

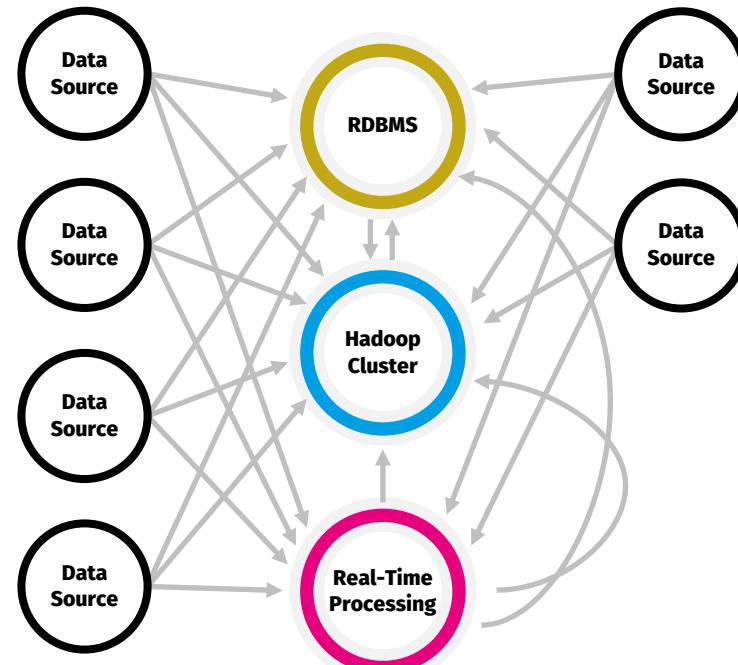
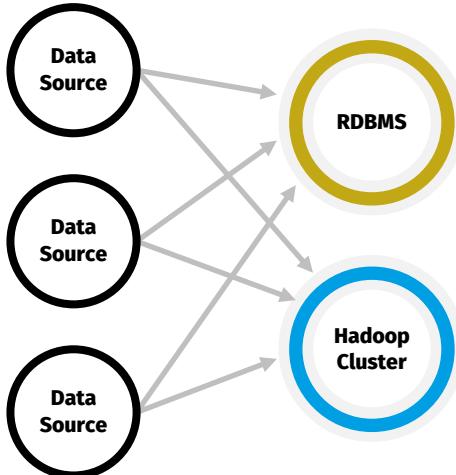
Die Welt der Daten



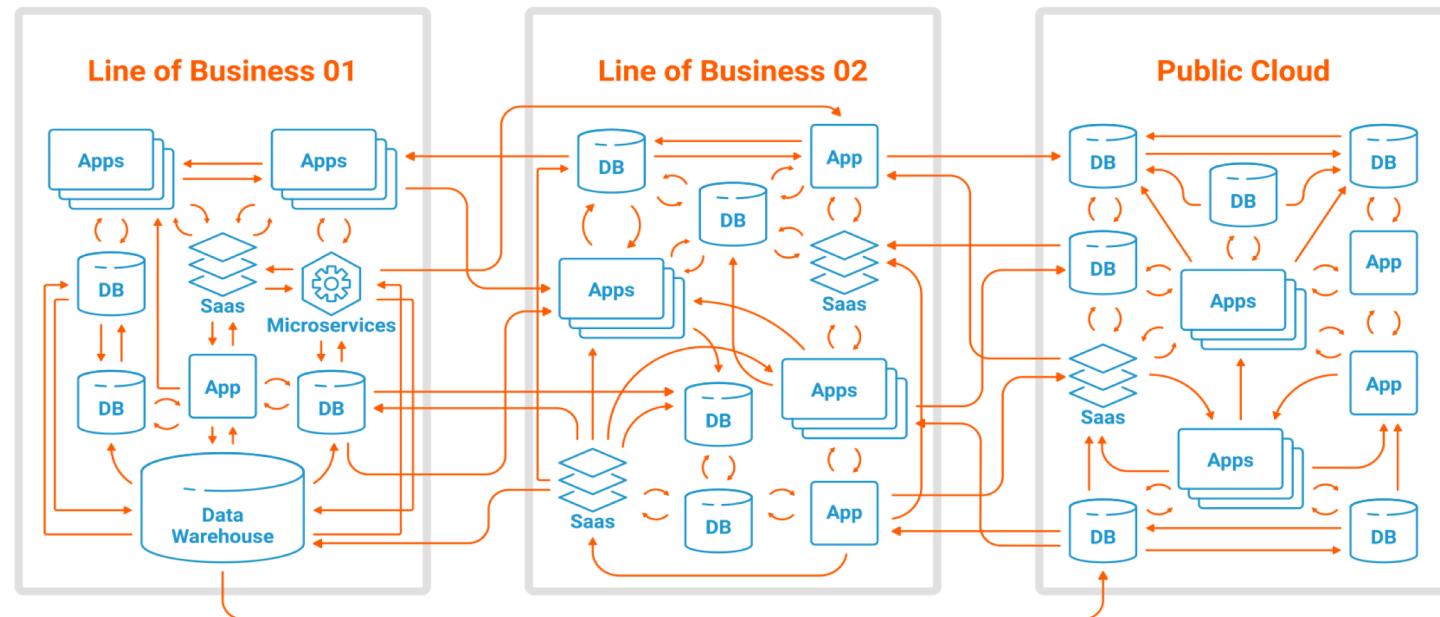
**Zuverlässige Informationen
sind unbedingt nötig für das
Gelingen eines Unternehmens.**

Christoph Kolumbus (1451 - 1506)

Die Welt der Daten



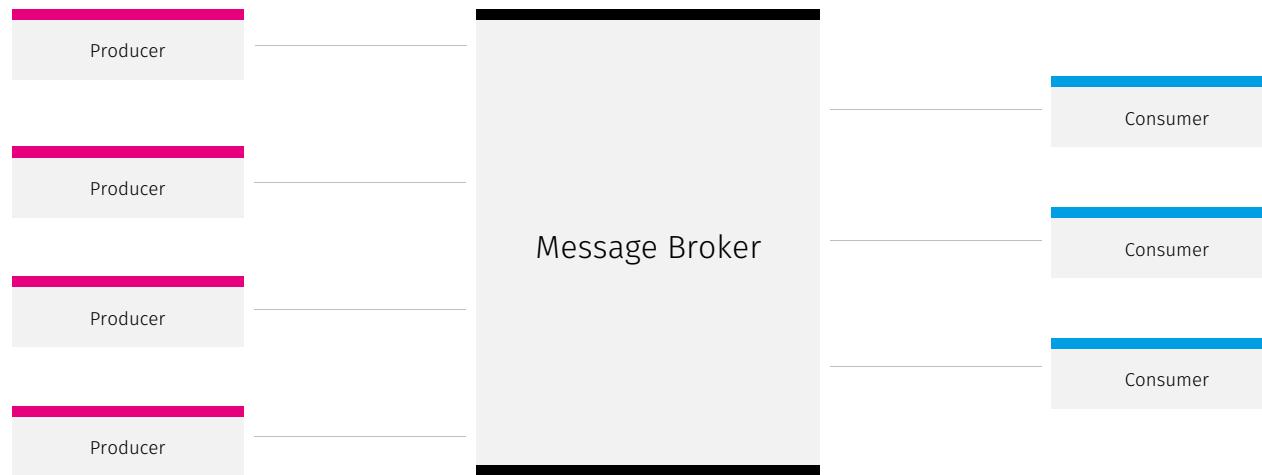
Die Welt der Daten



[Bildquelle: © confluent.io]

Message Broker/Kafka Basics

Message Broker



Message Broker

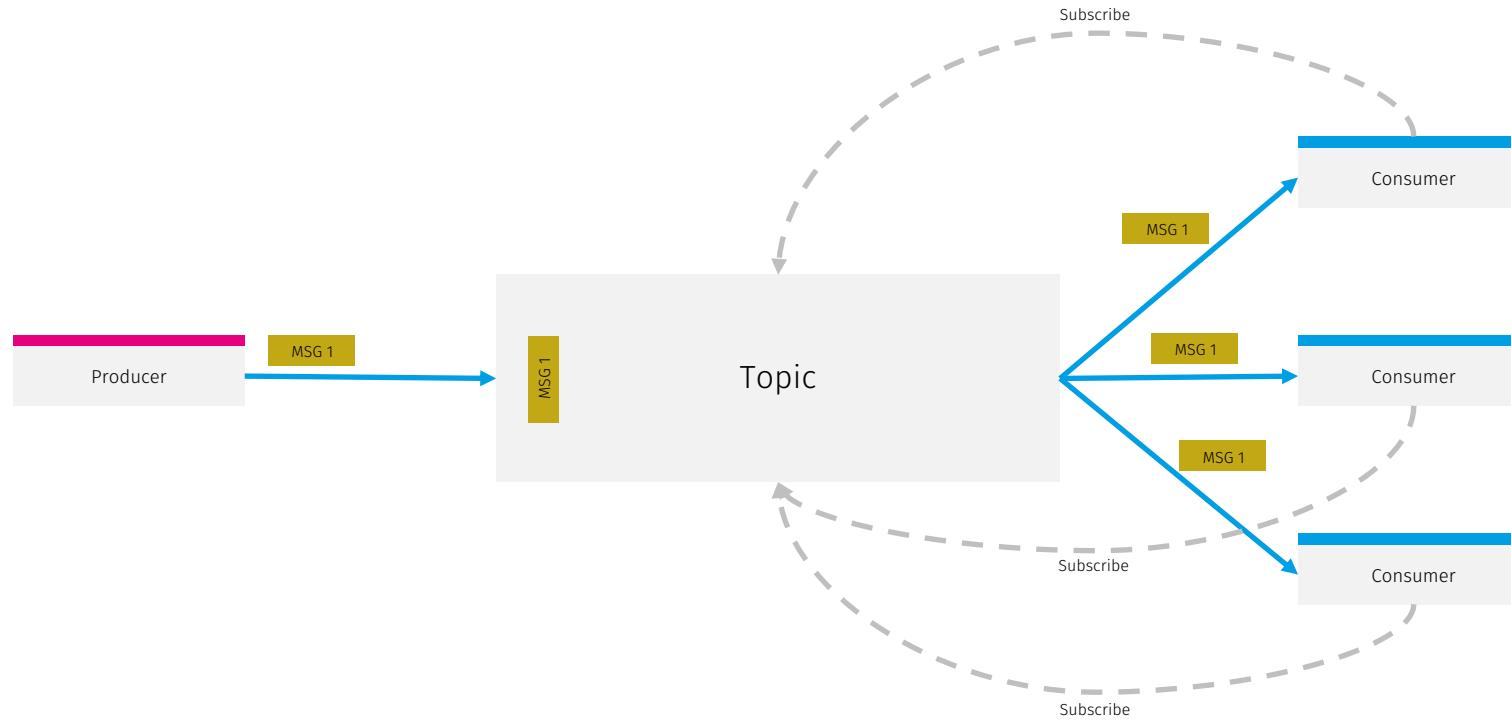
- Middle-Ware zum Empfang und Versand von Nachrichten
- Dient als Datenpipeline
- Lose Kopplung der Systeme

- Verarbeitung unterschiedlichster Arten von Nachrichten
 - Informationen von Sensoren
 - Transaktionen vom Börsenhandel
 - Bestellungen eines Shop-Systems
 - Transaction-Log einer Datenbank
 - Logfiles einer Webanwendung
 - etc.

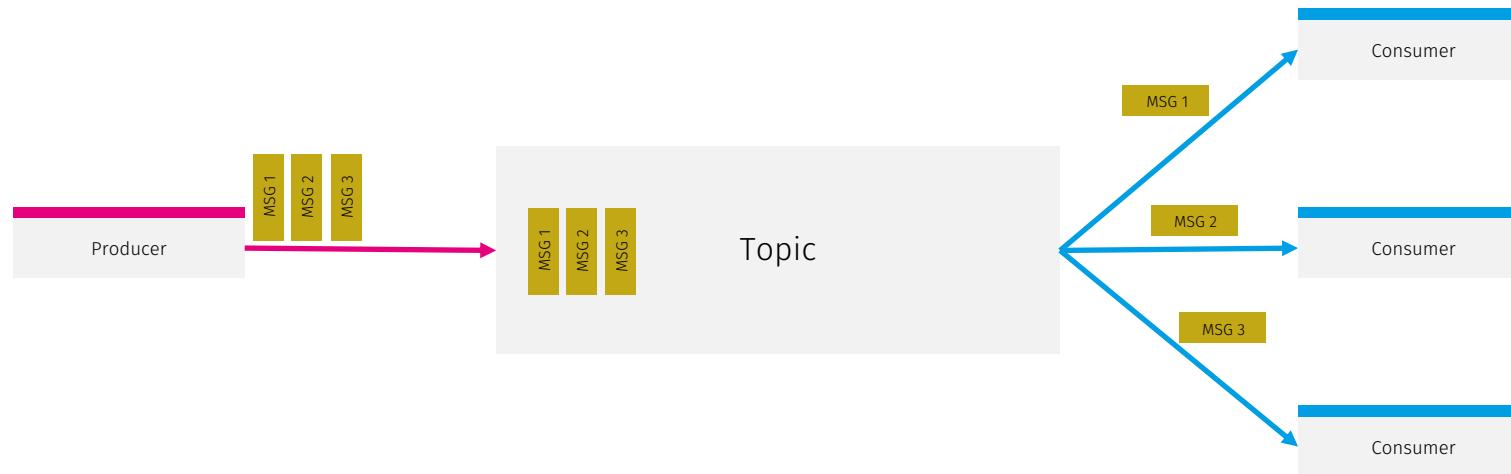
- Kommunikation: Publish-Subscribe oder Point-To-Point



Publish-Subscribe



Point-To-Point



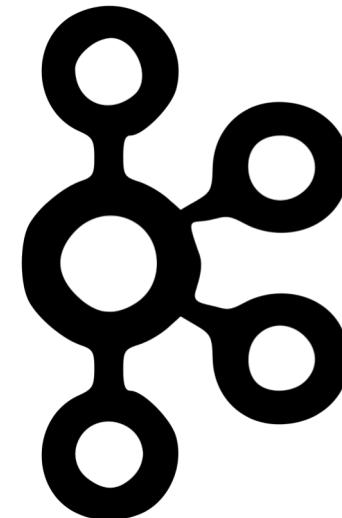
Warum Kafka

Vorteile

- Sehr hoher Datendurchsatz
- Exzellente Skalierbarkeit
- Redundanz
- Vorhaltung der Nachrichten
- Unterstützung Batch- als auch Real-Time-Verarbeitung der Messages

Abgrenzung zu anderen Message Brokern

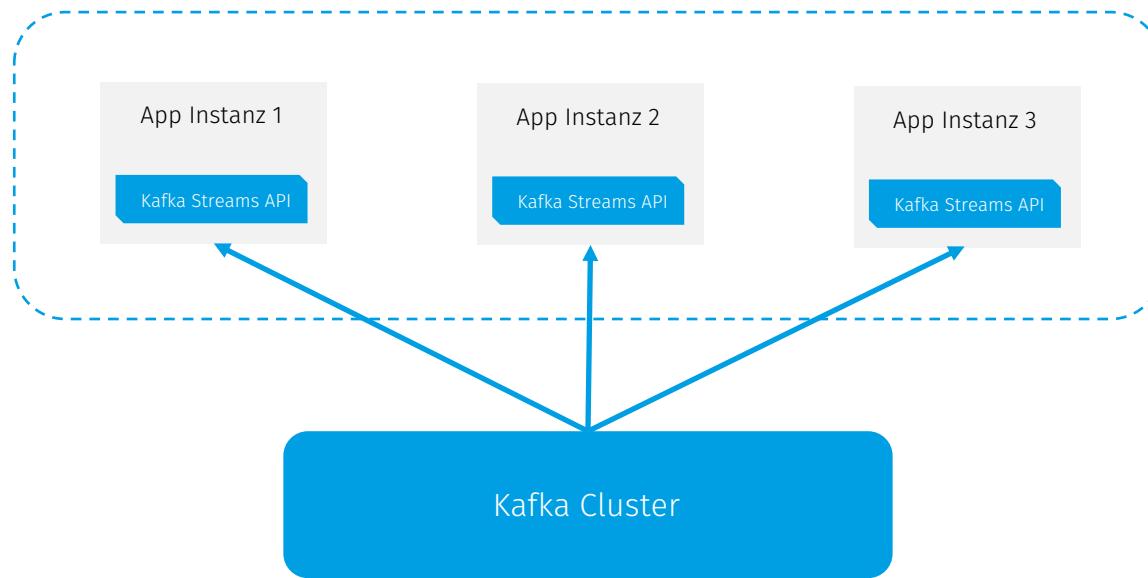
- Apache Kafka wurde von Anfang an für den Cluster-Betrieb entwickelt
- Nachrichten werden nicht gelöscht, wenn sie zugestellt wurden
- Nachrichten sind wieder abspielbar



Kafka Streams

- Nachrichten werden nicht nur von Producer zu Consumer weitergereicht
- On-The-Fly-Weiterverarbeitung von Nachrichten
- Ergebnisse des Streamings werden in Topics geschrieben
- Stream API ist eine eigenständige Java-Library

Kafka Streams

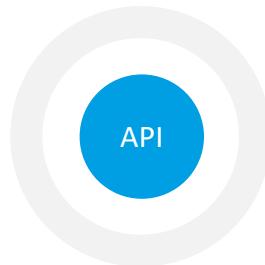


Kafka Streams



Kafka Stream DSL (Domain Specific Language)

- <https://docs.confluent.io/currentstreams/developer-guide/dsl-api.html>
- Empfohlen für die meisten Anwendungsfälle
- Business-Logik kann oft mit ein paar Zeilen Code abgebildet werden
- Kapselung der Stream-Processing-Komplexität



Kafka Processor API

- Der Low-Level Weg, um Stream-Processing mit Topologien aufzubauen
- Bietet mehr Flexibilität als Kafka Stream DSL



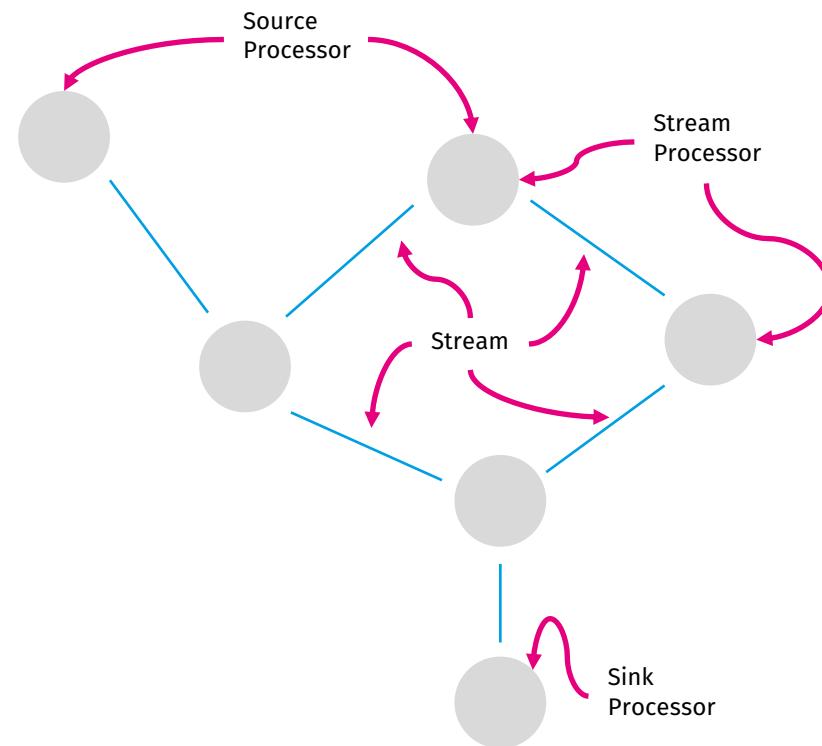
KSQL

- Streaming mittels SQL-ähnlicher Syntax

Stream Processing in der Theorie

Prozessor Topologien

- Stream Processors
- Source Processors
- Sink Processors



Einsatzgebiete von Streaming



IoT Bereich



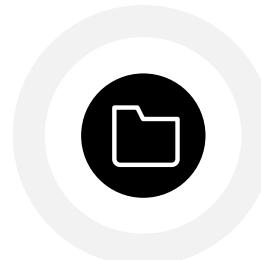
Finanzwelt

- Aktienhandel



Shop-Systeme

- Recommendations
- Page-Click-Tracking
- Logins

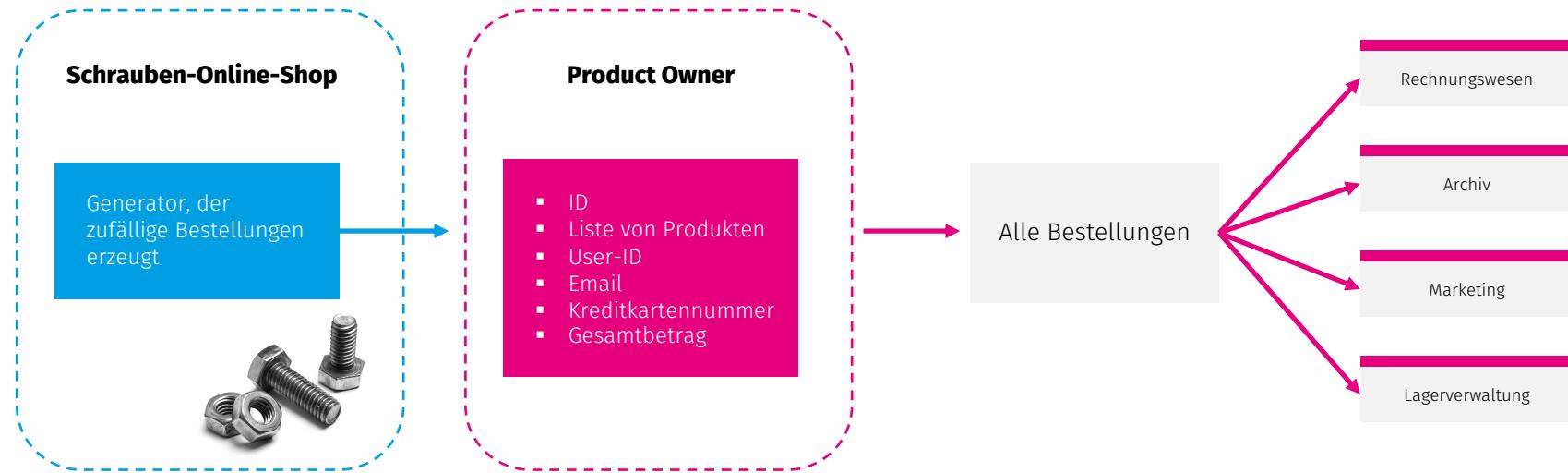


Logfile-Analysen

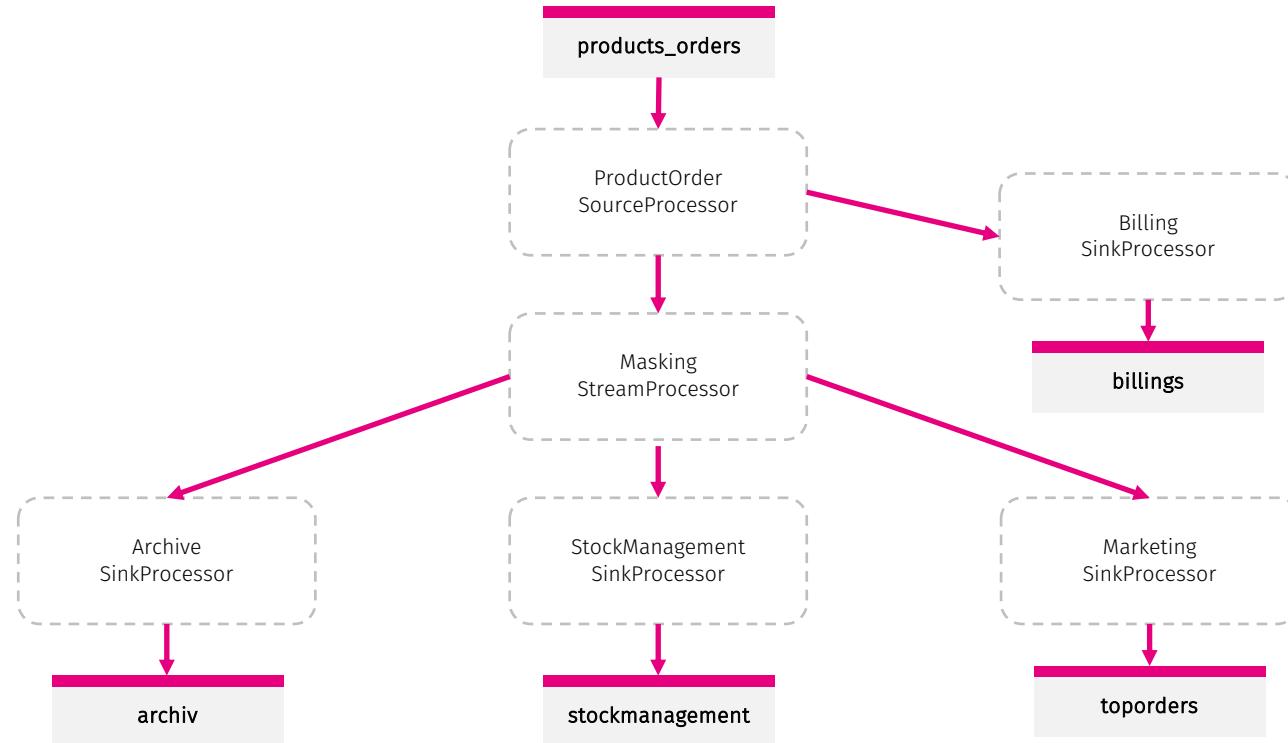
- Intrusion Detection

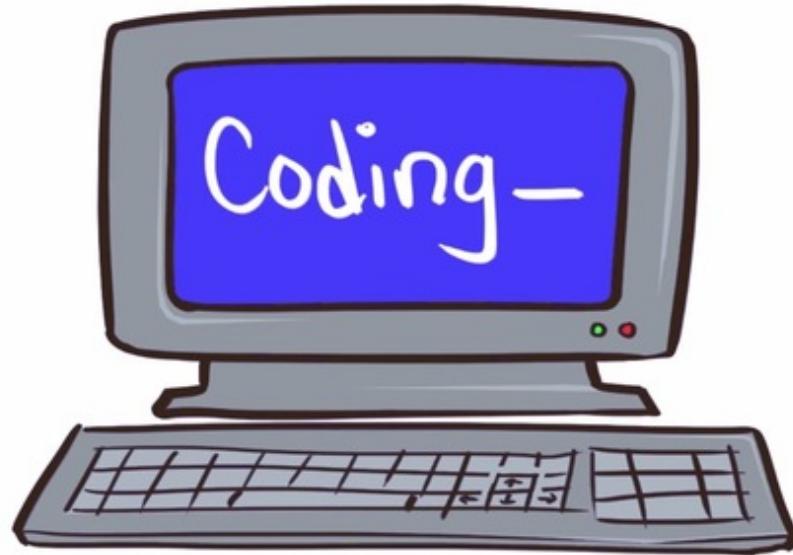
Showcase mit Kafka Streaming API

Showcase

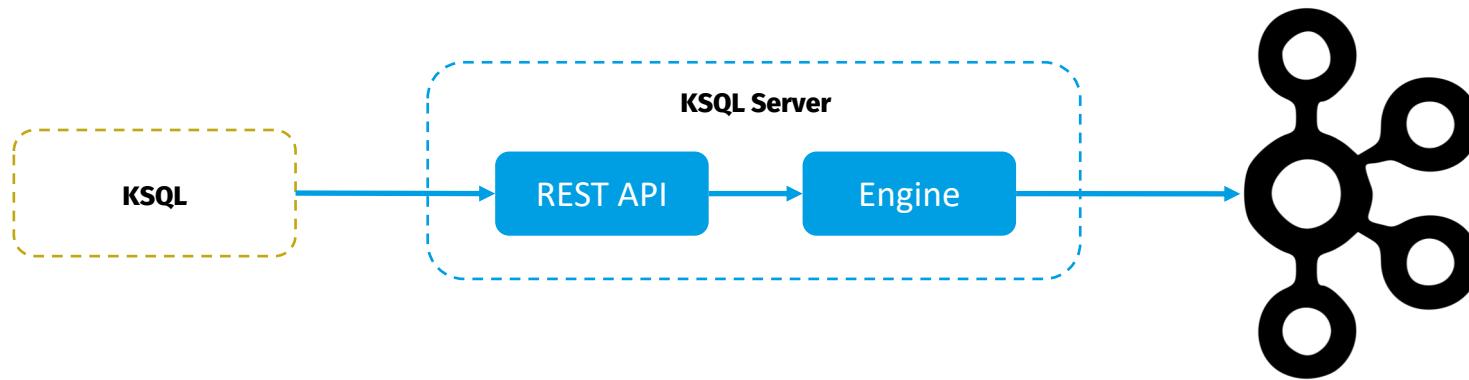


Showcase

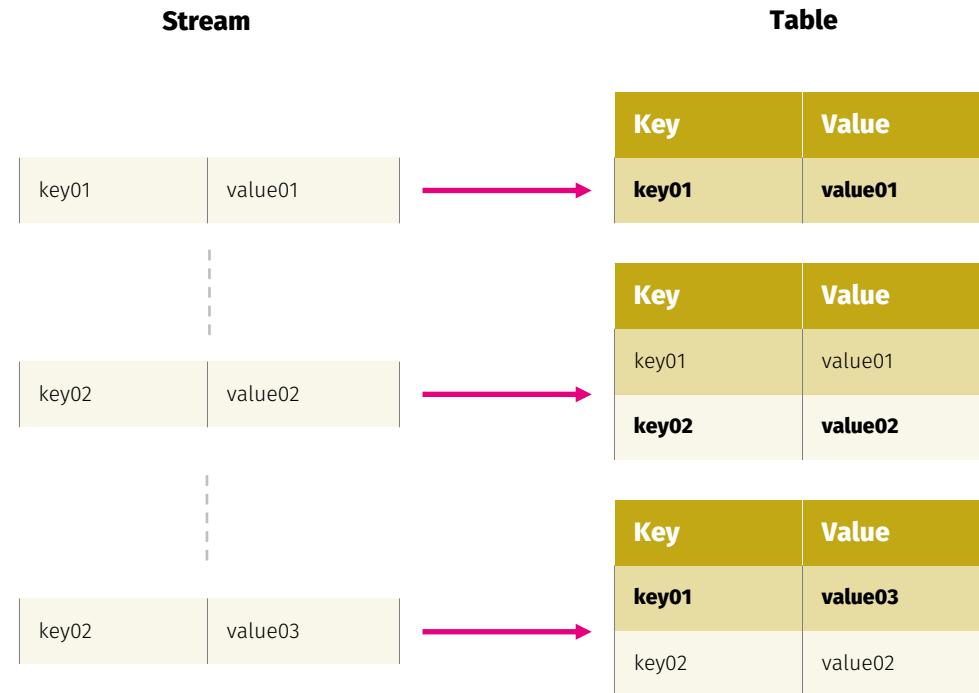




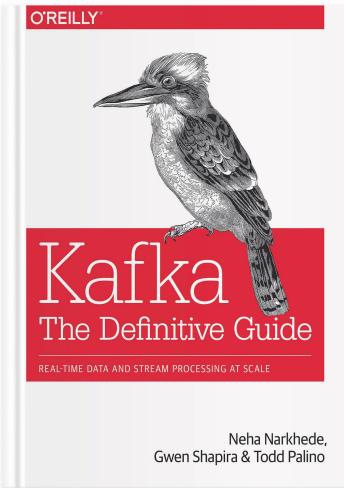
KSQL



Stream vs. Table



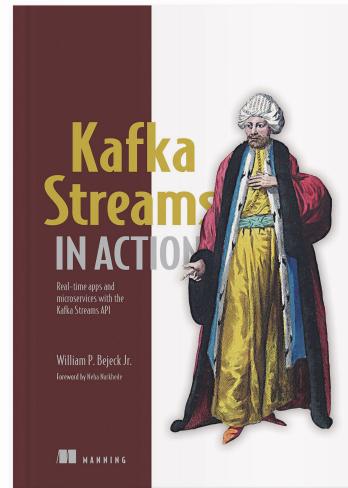
Buchtipps



Kafka: The Definitive Guide
Real-time data and stream processing at scale

von Neha Narkhede, Gwen Shapira,
Todd Palino

Neha Narkhede,
Gwen Shapira & Todd Palino



Kafka Streams in Action

von William P. Bejeck Jr.

Vielen Dank



<https://github.com/diva-e/kafka-streaming-talk.git>

Vielen Dank

<https://github.com/diva-e/kafka-streaming-talk.git>

